

Allosteric control in icosahedral capsid assembly

Guillermo R. Lazaro and Michael F. Hagan*

Martin Fisher School of Physics, Brandeis University, Waltham, MA, 02454

(Dated: April 28, 2016)

During the lifecycle of a virus, viral proteins and other components self-assemble to form an ordered protein shell called a capsid. This assembly process is subject to multiple competing constraints, including the need to form a thermostable shell while avoiding kinetic traps. It has been proposed that viral assembly satisfies these constraints through allosteric regulation, including the interconversion of capsid proteins among conformations with different propensities for assembly. In this article we use computational and theoretical modeling to explore how such allostery affects the assembly of icosahedral shells. We simulate assembly under a wide range of protein concentrations, protein binding affinities, and two different mechanisms of allosteric control. We find that, above a threshold strength of allosteric control, assembly becomes robust over a broad range of subunit binding affinities and concentrations, allowing the formation of highly thermostable capsids. Our results suggest that allostery can significantly shift the range of protein binding affinities that lead to successful assembly, and thus should be accounted for in models that are used to estimate interaction parameters from experimental data.

I. INTRODUCTION

The assembly of a virus outer protein shell (capsid) requires a delicate balance among thermodynamic and kinetic constraints. The proteins must assemble quickly to evade proteolysis and detection by the host, and their capsid must be sufficiently thermostable to survive intact under potentially harsh conditions while searching for a new infection target. Yet, self-assembly of ordered structures usually requires weak, reversible interactions among the components, since strong interactions lead to kinetic traps [1–3]. Many viruses must also control the time and place of assembly, so that the capsid can select specific components from amidst a crowded cellular environment. A number of strategies have been proposed by which viruses control their assembly process to ensure productive and timely capsid formation (*e.g.* [4–8]). One such strategy is allosteric regulation [8, 9], in which capsid proteins sample an ensemble of conformational states with different propensities for assembly, with the relative populations of different states influenced by binding of proteins or other molecules. In this article, we theoretically and computationally examine how allostery at the level of protein-protein interactions can lead to self-regulation of assembly kinetics.

Learning mechanistic information such as allosteric regulation from experiments alone is challenging because most assembly intermediates are transient, and thus not readily observed. For example, a wealth of information has been obtained from *in vitro* experiments in which capsid assembly kinetics are monitored by size exclusion chromatography (SEC) or x-ray or light scattering (*e.g.*) [10–17]. However, since intermediates are usually undetectable, these bulk techniques primarily report on the concentrations of assembled capsids and unassembled subunits. Recently, techniques which monitor individual capsids or can detect transient intermediates have begun to address this limitation [18–27]. However, even these techniques provide structural data at limited resolution and cannot characterize the full ensemble of intermediates.

Therefore, theoretical models are needed to obtain a complete understanding of capsid assembly from such experimental data.

Theoretical models have already played an important role in relating experimental data to assembly pathways and the driving forces that control them. For example, binding affinities have been estimated by fitting the ratio of assembled capsids to unassembled subunits measured at long times to the equilibrium law of mass action [28, 29]. Assembly kinetics have been analyzed using models in which capsid formation is viewed as the assembly of rigid subunits (lacking internal degrees of freedom) into polyhedral shells. These models can be formulated as a master equation and solved numerically (*e.g.* [11, 30–33]) or analyzed by stochastically generating trajectories consistent with the master equation (*e.g.* [34–40]). Despite simplifications used to make these methods tractable, they capture many features of experimental assembly kinetics. Fitting their results against light scattering data has enabled estimates of physical parameters such as subunit-subunit binding affinities and rate constants [11, 14, 41–43]. These results suggest that capsid subunit-subunit binding affinities are generically weak, on the order of 4 kcal/mol [3, 28]. When binding affinities exceed this limit, assembly is limited by kinetic traps, in which the formation of long-lived disordered or partially assembled structures inhibit capsid formation.

Despite these important insights, current models do not capture all aspects of experimental data. Typically, models cannot quantitatively reproduce kinetics at all timescales across a range of concentrations [41]. Similarly, capsid disassembly exhibits a surprising degree of hysteresis considering the measured weakness of subunit binding affinities [44, 45]. For example, assembled capsids are highly stable in infinite dilution; *e.g.* hepatitis B virus (HBV) capsids exhibit virtually no subunit dissociation even on a timescale of months [46]. While this observation could be accounted for by a post-assembly maturation process that increases capsid stability, as yet there is no evidence for this in HBV.

One feature which has not yet been incorporated in most assembly models is that capsid proteins sample multiple conformational states, with different propensities for assembly. For example, structural studies on HBV [9], brome mosaic virus

* hagan@brandeis.edu

(BMV) and HIV [47] find that their capsid proteins adopt conformations in solution that are incompatible with insertion into a capsid, suggesting that the protein’s primary conformation in solution is inactive for assembly (assembly-inactive), and that capsid formation requires a transition to an assembly-active conformation. These observations have led to the hypothesis that assembly of icosahedral viruses may be subject to allosteric regulation [9, 28, 41, 47–50].

It has long been appreciated that protein conformational switching can play a key role in the kinetics of assembling filamentous structures. Asakura [51] showed that the elongation rate of the bacterial flagellum is limited by the rate of flagellin monomer undergoing a conformational transition to its bound form. From this observation, he deduced that interaction with the flagellum triggers the conformational transition in the monomer by an ‘induced fit’. Similar observations were made for other filamentous assemblies, including the bacteriophage T4 tail [52] and tobacco mosaic virus [53–56]. Caspar coined the term ‘autostery’ [8, 57] to describe such an induced fit process, in which a protein that exists in an assembly-inactive conformation in solution is driven to switch to an assembly-active conformation by interacting with other copies of itself within an assemblage. By controlling the rate of nucleation in comparison to elongation, autostery could provide mechanism for self-regulation of assembly.

The role of an assembly-active/inactive transition in the assembly kinetics of icosahedral viruses has received less attention; more work has focused on the roles of conformational switching in overcoming the geometrical constraints imposed by an icosahedral geometry [58–61] and structural polymorphism [19, 62–65]. Although Packianathan et al. [9] was primarily an experimental study, they also used a rate equation approach to compare the behaviors of assembly with no allosteric regulation, with induced fit (autostery), or allostery without induced fit. They concluded that the induced fit mechanism leads to productive assembly and could increase hysteresis associated with disassembly, but that allostery without induced fit did not lead to productive assembly, because such strong subunit affinities were required that kinetic trapping resulted.

In this article we perform a more extensive theoretical and computational investigation of the effects of allostery on assembly kinetics and their sensitivity to kinetic trapping. In contrast to Ref. [9], we find that both mechanisms of allosteric regulation (with or without induced fit) can drive productive assembly, although induced fit allows for productive assembly over a wider parameter range. Under moderate parameter values, allostery does not enhance assembly robustness — the width of the range of subunit concentrations or binding affinities leading to productive assembly is not increased. However, under sufficiently strong allosteric control (meaning that the population of unassembled subunits is strongly shifted toward the assembly-inactive conformation), high assembly yields are achieved over a broad range of parameters, including high binding affinities. Our results highlight the importance of accounting for allostery in models used to estimate parameters from experimental data.

This paper is organized as follows. In the next section, we

incorporate allostery into a computational model and a master equation description of assembly. In section ‘Scaling estimates for the effect of allostery on assembly timescales’ we develop scaling estimates for the effects of allostery on assembly timescales and assembly robustness. Then, in the Results we test these scaling estimates against numerical results from the computational and master equation models. In the Discussion we summarize the key observations, identify potential further extensions to the models, and discuss implications for estimating parameters by fitting against experimental data. Finally, the appendix provides further details about the models and compares the computational and master equation models.

II. MODELS

A. Allostery Models

To represent allostery in our models, unassembled subunits interconvert between assembly-inactive and active conformations, with equilibrium constant $K_A = \exp(-g_A/k_B T)$ and $g_A > 0$ the unfavorable free energy associated with a monomer adopting the active state. The case of no allostery corresponds to the limit $K_A \rightarrow \infty$. For simplicity, here and throughout, we focus on the limit where the active/inactive interconversion rate is fast relative to assembly timescales. We discuss effects of a slow conformational interconversion step in the Discussion.

The two allostery behaviors that we consider throughout the paper are illustrated by the schematics in Fig. 1. In the first, there is allostery but no induced fit, meaning that only subunits in the active conformation can associate with any size intermediate. In the second, subunits in the inactive conformation can associate with intermediates equal to or larger than a nucleus comprising n_{nuc} subunits. For notational consistency, we refer to these cases respectively as ‘No Induced Fit’ and ‘Induced Fit’, and the case without allostery (all subunits are active) as ‘No Allostery’.

Structural evidence suggests that the active-inactive transition in the HBV protein is closely linked to formation of a stable critical nucleus [9]. The latter is associated with the geometry of an icosahedral shell. Since smaller intermediates have fewer interactions per subunit than large ones, stability increases with intermediate size. The critical nucleus refers to the smallest intermediate (or ensemble of intermediates) from which assembly into a complete capsid is more probable than complete dissociation. In vitro experiments on several viruses have determined critical nucleus sizes corresponding to small polygons (*e.g.* a pentamer or trimer of dimers) [2]. Similarly, in the computational model described below we find that under certain conditions the critical nucleus corresponds to the smallest polygon that can form, which is a pentagon as shown in Fig. 2C (see also Refs. [66, 67]). Therefore, we consider the critical nucleus as the minimum seed capable of driving a conformation change. I.e., inactive subunits can bind to an intermediate at or above the critical nucleus size, but only active subunits can associate to pre-critical intermediates (Fig. 1).

We now describe the two modeling approaches we use to

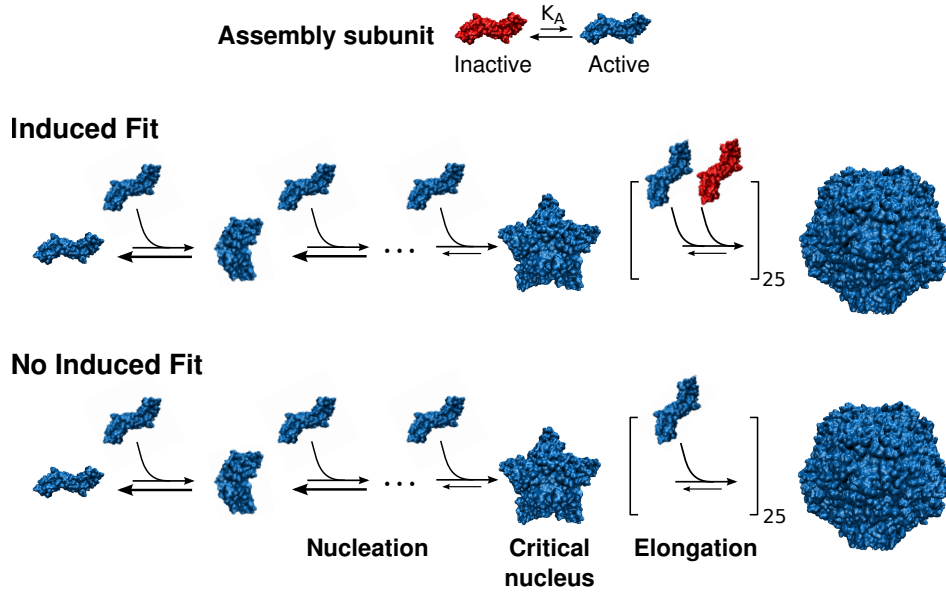


FIG. 1. The two mechanisms of allostery that we consider in this article. The irreducible assembly unit (a protein dimer in this schematic) inter-converts between assembly-inactive and assembly-active conformations, with equilibrium constant K_A . In both mechanisms, only assembly-active subunits can combine to form intermediates smaller than a critical nucleus. In the ‘Induced Fit’ mechanism, both assembly-active and assembly-inactive subunits can bind to larger intermediates, whereas in the ‘No Induced Fit’ mechanism, only assembly-active subunits can bind throughout the assembly process.

study the effects of allostery on assembly.

B. Computational model

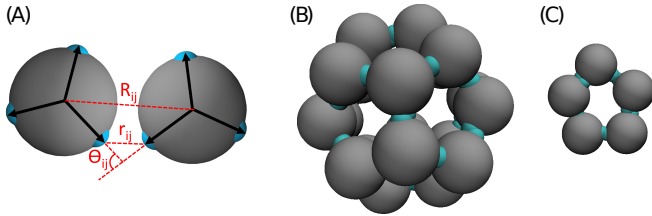


FIG. 2. Computational model geometry. (A) Geometry of two interacting subunits with the bond vectors depicted as arrows and attractors colored teal. The angle between each of the subunit bond vectors is 108° , and the interactions are described in the Appendix (Eqs. (16)). The dihedral angle ϕ_{ij}^b is not shown. (B) The complete capsid, which contains 20 subunits arranged with icosahedral symmetry. (C) Critical nucleus for binding of assembly-inactive subunits in the ‘Induced Fit’ model.

We consider a model for the assembly of icosahedral shells used in Ref. [68], in which subunits are modeled as rigid bodies, with excluded volume interactions represented by spherically symmetric repulsive forces, and the complementary subunit-subunit interactions that drive assembly represented by directional attractions. The lowest energy states in the model correspond to ‘capsids’, which consist of $N = 20$ sub-

units (each of which could represent a protein trimer) in a shell with icosahedral symmetry. Because the spatial positions and orientations of all subunits are explicitly tracked, there are no assumptions about assembly pathways or the structures that emerge from assembly.

Following the approach developed by Schwartz et al. [60], the subunit excluded volume is spherically symmetric and three attractive patches (bond vectors) are rigidly fixed to the subunit, with each pair of bond vectors forming an angle of 108° (see Fig. 2 and Eq. (16)). There is a favorable interaction between subunits when (1) the ends of bond vectors nearly overlap, (2) the bond vectors are nearly anti-parallel, and (3) the secondary bond vectors are nearly coplanar. Twenty subunits realizing these conditions results in the minimum energy target structure (a complete capsid) shown in Fig. 2. The interaction strength is tuned by the parameter ϵ_b . All interaction potentials are pairwise.

The subunit-subunit interaction follows additional rules depending on the type of allostery being modeled. For ‘No Allostery’, all pairs of subunits meeting the interaction criteria listed above experience attractive interactions. For ‘No Induced Fit’, subunits stochastically switch between inactive and active conformations. Only pairs of active subunits satisfying the binding criteria experience attractive interactions. Since there is no autostery in this case, the effects of allostery on assembly arise only due to the cooperativity of capsid assembly. For ‘Induced Fit’, we define the threshold for autostery activity to be the formation of at least one polygon (any closed cycle of subunit-subunit interactions, such as a pentamer). Thus, in this case interactions are possible for any

pair of active subunits, and also between inactive subunits and subunits within a partial capsid which has at least one complete polygon. Full details of the model are given in the appendix.

Simulation parameters. The parameters of the model are the energy associated with the attractive potential, ϵ_b , and the specificity of the directional attractions, which is controlled by the angular parameters θ_c and ϕ_c . Subunit positions and orientations were propagated using overdamped Brownian dynamics according to a second order predictor-corrector algorithm [69, 70], with the unit of time $t_0 = \sigma^2/D$, where D is the subunit diffusion coefficient and σ is the subunit diameter. We simulated systems with 500 subunits in a periodic box with side length 17, where all distances are measured in units of the subunit diameter σ . For each parameter set, results were averaged over 20 or more independent simulations. The orientational specificity parameters were $\theta_c = 0.5$ and $\phi_c = \pi$. These parameters tend to disfavor the formation of incorrect subunit-subunit interaction geometries and thus inhibit formation of malformed capsids, allowing us to study the high affinity limit. However, such configurations do arise at higher binding energies as discussed below. To obtain dimensionless units, we rescale energies by $k_B T$ and times by t_0 . The subunit conformational switching rate was $2.5/t_0$. Simulations were initialized by generating random positions and orientations for subunits, with subunit positions that led to subunit-subunit overlap (positive potential energies in excess of $1 k_B T$) rejected.

C. Master equation model for capsid assembly

We also consider a master equation description of polyhedral shell assembly, which is sufficiently computationally tractable to allow modeling assembly over broad parameter ranges. Specifically, we extend the ‘nucleation and growth’ model described in Ref.[66] to include multiple subunit conformations. This model was based on the work of Zlotnick and coworkers [11, 30, 31], and consists of a system of coupled rate equations that describe the time evolution of concentrations of empty capsid intermediates:

$$\begin{aligned} \frac{dc_1}{dt} &= -k_1 c_1^2 + 2\bar{k}_2 c_2 + \sum_{n=2}^N -k_n c_n c_1 + \bar{k}_n c_n \\ \frac{dc_n}{dt} &= k_{n-1} c_1 c_{n-1} - k_n c_1 c_n - \bar{k}_n c_n + \bar{k}_{n+1} c_{n+1} \quad n = 2 \dots N \end{aligned} \quad (1)$$

where c_n is the concentration of intermediates with n subunits, and k_n and \bar{k}_n are respectively association and dissociation rate constants for intermediate n . The initial condition is $c_1(0) = c_0$, $c_n(0) = 0$ for $n > 1$. The extensions of Eq. (1) to describe allostery are given in the appendix.

There are several important assumptions underlying the master equation: malformed capsids are not considered [60, 62–64, 71–74], assembly proceeds along a single pathway [34, 40, 75], only single subunits can bind or unbind, and only one k_n and \bar{k}_n are considered for each size n (averaged over

all intermediates of that size). However, these assumptions are not present in the BD simulations described above, and we find close agreement between the two approaches (see appendix Fig. 11). Moreover, Ref. [66] showed that extending Eqs. (1) to relax these simplifications does not qualitatively change their predictions. Most importantly, rate equations of this form capture many features of experimental assembly kinetics data (e.g. [11, 76]).

The association and dissociation rate constants are related by detailed balance $\bar{k}_n = k_n \exp(\Delta g_n / k_B T) / v_0$, with $\Delta g_n = G_n - G_{n-1}$ the change in free energy due to association of a subunit. Association free energies Δg_n , which can be fit to experimental data using the law of mass action [28, 77, 78], include hydrophobic and electrostatic interactions [79] and depend on pH and salt concentration [28]. Specifying the assembly model requires defining the set of intermediates n and the transition rates $\{k_n, \bar{k}_n\}$.

The model we use here is based on those of Zlotnick and coworkers [11, 30, 31], in which the subunit-subunit association free energy for intermediate n is proportional to the number of new subunit-subunit contacts $n_{c,n}$ formed by addition of a subunit to that intermediate [80]. Specifying $\{\Delta g_n\}$ thus requires defining the geometry of each intermediate. This usually begins with specifying the geometry of a capsid and its subunits in terms of a polyhedron (for example, see Fig. 2 or Fig. 1 in Ref. [31]), and assuming that assembly proceeds along a single path. The path can be comprised of the lowest energy intermediate for each size n [30] or correspond to an ‘average’ pathway [31] in which all subunits, except during the initial and final stages of assembly, make the same average number of contacts. We choose the latter definition, since it is simpler and both definitions lead to a qualitatively similar behavior. Specifically, the association rate constant k is independent of intermediate size and association free energies are given by $\Delta g_n = g_b$ before nucleation ($n < n_{\text{nuc}}$) and $\Delta g_n = g_{\text{elong}}$ during elongation ($n_{\text{nuc}} \leq n < N - 1$), where n_{nuc} is the critical nucleus size. Finally, inserting the last subunit makes the maximum number of possible contacts and thus enjoys the most favorable association free energy g_N .

Because this model was previously explored extensively in the absence of conformation changes [66], we focus on one set of interaction parameters (except where noted otherwise): capsid size $N = 120$ corresponding to 120 dimer subunits in hepatitis B virus [28], critical nucleus size $n_{\text{nuc}} = 5$ (a pentamer of dimers), association rate constant $k = 10^5 \text{ M}^{-1} \text{ s}^{-1}$ [76], and free energy parameters $g_b = 7k_B T$ ($\approx 4 \text{ kcal/mol}$) [28], $g_{\text{elong}} = 2g_b$ and $g_N = 2g_{\text{elong}}$, which imply that adding a subunit becomes on average twice as favorable after nucleation and four times as favorable for the final subunit.

III. SCALING ESTIMATES FOR THE EFFECT OF ALLOSTERY ON ASSEMBLY TIMESCALES

To gain an intuitive understanding of how allostery can affect assembly, in this section we derive simple scaling estimates for the timescales associated with the two assembly mechanisms shown in Fig. 1, based on the master equation

model (Eqs. 1). Although we introduce a number of simplifications, in the next section we show that the resulting scaling estimates apply at least qualitatively when these simplifications are relaxed in the computational and theoretical models. We closely follow Ref. [66], except that we extend the analysis to include allostery.

We consider a system of capsid protein subunits with total concentration c_0 that assemble into capsids with N subunits. The word subunit refers to the basic assembly unit, which could be a protein dimer or larger oligomer [2]. As in the master equation model, we break the assembly of a capsid into ‘nucleation’ and ‘elongation’ phases. For simplicity we assume that the association rate constant k is independent of intermediate size, so that for the ‘No Allostery’ reaction rates of association to each intermediate are kc_1 with c_1 the concentration of free subunits. We assume the limit of fast conformational interconversion, so for the ‘No Induced Fit’ case, association rates are given by $k f_A c_1$ with $f_A = K_A / (1 + K_A)$. For ‘Induced Fit’, association rates are $k f_A c_1$ for intermediate size $n < n_{\text{nuc}}$ and kc_1 for $n \geq n_{\text{nuc}}$.

We now write the time required for an individual capsid to assemble as $\tau = \tau_{\text{nuc}} + \tau_{\text{elong}}$ with τ_{nuc} and τ_{elong} the average times for nucleation and growth, respectively. The elongation timescale can be estimated by the mean first passage time for a biased random walk with a reflecting boundary conditions at n_{nuc} and absorbing boundary conditions at N , with forward and reverse hopping rates given by the subunit association and dissociation rates respectively [66]. For early in the reaction, when $c_1 \approx c_0$, this results in

$$\tau_{\text{elong}} \cong N / k c_0 f_A^{m_A} \quad (2)$$

where $m_A = 1$ for ‘No Induced Fit’ and $m_A = 0$ for ‘Induced Fit’. We see that the elongation time is equal for ‘No Allostery’ and ‘Induced Fit’ since in the latter case all subunits can bind to post-nucleated intermediates. Ref. [66] showed that the duration of the lag phase in light scattering is proportional to τ_{elong} , thus predicting that the lag time should scale inversely with subunit concentration. This prediction was recently confirmed by experiments on HBV assembly [81].

The mean nucleation time at the beginning of the reaction can be estimated from the statistics of a random walk biased toward disassembly [66]. Including conformation dynamics results in $\tau_{\text{nuc}}^{\text{min}} \approx k^{-1} \exp(G_{\hat{n}}/k_B T) c_0^{-n_{\text{nuc}}} f_A^{-n_{\text{nuc}}}$, where $\hat{n} = n_{\text{nuc}} - 1$ so that $G_{\hat{n}}$ is the interaction free energy of the structure just below the critical nucleus. This estimate can be understood by noting that the pre-critical nucleus is present with concentration $c_{\hat{n}} \cong \exp(G_{\hat{n}}) c_0^{\hat{n}} f_A^{\hat{n}}$, and the rate of active subunits associating to the precritical nucleus is given by $k c_0 f_A$.

However, because free subunits are depleted by assembly, the nucleation rate never reaches this value, and net nucleation asymptotically approaches zero as the concentration of completed capsids approaches its equilibrium value. Thus, we estimate the median assembly time $\tau_{1/2}$ (the time at which the reaction is 50% complete) by treating the system as a two-state reaction with n_{nuc} -th order kinetics, which yields [66]

$$\tau_{1/2} \cong \frac{A_{1/2} P_N}{N k} \exp(G_{\hat{n}}/k_B T) c_0^{-\hat{n}} f_A^{-\hat{n}-1} \quad (3)$$

with $A_{1/2} = \frac{2^{\hat{n}-1}}{\hat{n}}$, and P_N as the equilibrium fraction of subunits in complete capsids. The factor of N^{-1} in Eq. 3 accounts for the fact that N subunits are depleted by each assembled capsid.

When capsid growth is fast compared to nucleation, the expressions Eq. 2 and Eq. 3 respectively predict the duration of the lag phase and the median assembly time. However, these relations begin to fail as intermediate concentrations build up above a crossover concentration c_c at which the initial elongation and nucleation times are equal

$$c_c \cong N^{-\frac{2}{\hat{n}}} e^{\frac{G_{\hat{n}}}{\hat{n} k_B T}} f_A^{-\frac{\hat{n}-m_A+1}{\hat{n}}} \quad (4)$$

Significant kinetic trapping then sets in above a threshold concentration set by $\tau_{\text{elong}} = \tau_{1/2}$:

$$c_{\text{kt}} \cong (A_{1/2} N^{-2})^{\frac{1}{\hat{n}-1}} e^{\frac{G_{\hat{n}}}{(\hat{n}-1) k_B T}} f_A^{-\frac{\hat{n}-m_A+1}{\hat{n}-1}}. \quad (5)$$

While the above analysis identifies a maximum concentration above which the reaction will become kinetically trapped, we can also identify a minimum concentration below which the median assembly time becomes longer than the maximum timescale of the experiment or simulation τ_{max}

$$c_{\text{min}} \cong (A_{1/2} / \tau_{\text{max}} N k)^{-\frac{1}{\hat{n}}} e^{\beta G_{\hat{n}} / \hat{n}} f_A^{-\frac{\hat{n}+1}{\hat{n}}} \quad (6)$$

which applies to both ‘No Induced Fit’ and ‘Induced Fit’.

A similar analysis can be performed for fixed subunit concentration and varying g_b , corresponding to changing pH or salt concentration. For example, if we assume that pre-critical intermediates with \hat{n} subunits have $\hat{n} - 1$ subunit-subunit interactions (Fig. 1), then $G_{\hat{n}} = (\hat{n} - 1) g_b$ and the minimum affinity for assembly in finite time τ_{max} is shifted according to

$$g_b^{\text{shift}} = g_b + \frac{\hat{n} + 1}{\hat{n} - 1} \log f_A. \quad (7)$$

A. The influence of conformation dynamics on assembly robustness

Based on the above scaling estimates, we now examine whether introducing allostery makes assembly more robust — *i.e.*, at a finite timescale τ_{max} relevant to an experiment or a cell, does allostery increase the range of concentrations over which productive assembly occurs? Specifically, we assume that kinetic trapping precludes productive assembly within τ_{max} , and consider the ratio between the minimum concentration leading to significant nucleation, and the maximum concentration leading to kinetic trapping, $\delta c = c_{\text{kt}} / c_{\text{min}}$. Since nucleation with allostery must occur within τ_{max} , we increase the binding affinity according to Eq. (7), resulting in

$$\delta c \cong \exp(\beta g_b^{\text{shift}} / \hat{n}) f_A^{\frac{m_A}{\hat{n}-1}}. \quad (8)$$

Since $m_A = 0$ for ‘Induced Fit’, allostery with induced fit has no effect on the range of concentrations over which assembly occurs (for moderate interaction strengths), while in the absence of induced fit, allostery renders assembly kinetics more

sensitive to concentration (since $f_A < 1$). However, either allosteric mechanism can significantly increase the maximum thermostability of a capsid that can be achieved without kinetic trapping. For a fixed concentration c_0 , equating Eqs. (2) and (3) shows that the minimum binding free energy g_b^{kt} below which kinetic trapping occurs is decreased (higher binding affinity) according to

$$g_b^{\text{kt}} = g_b^{\text{kt}}(f_A = 1) + \frac{\hat{n} + 1 - m_A}{\hat{n} - 1} \log f_A. \quad (9)$$

The maximum kinetically accessible thermostability of a complete capsid is then controlled by the free energy per subunit, $g_{\text{sub}} = n_B/2g_b^{\text{kt}}$ with n_B the number of contacts each subunit makes with its neighbors. For the case we consider below with $n_{\text{nuc}} = 5$ and $n_B = 4$, g_{sub} increases by a factor $n_B m_c / g_A / 2 \approx 3.33 g_A$. Thus, even a modest activation energy could substantially increase the capsid stability. This effect, together with the asymptotic approach to equilibrium an assembly reaction discussed below, could contribute to observations of unexpectedly large hysteresis between capsid assembly and disassembly.

Strong allostery drives robust assembly in the high-affinity limit. While the above analysis shows that allostery does not lead to more robust assembly for moderate interaction strengths, we observe dramatically different behavior with strong interactions, $g_b \gg k_B T$. It is well-known that unregulated assembly fails due to kinetic trapping in this limit. Since subunit-subunit interactions are effectively irreversible, the effective critical nucleus size is reduced to a dimer regardless of the capsid geometry, and elongation becomes slower than nucleation for any feasible parameters. However, equating Eqs. (2) and (3) shows that this trap can be avoided by a sufficiently high g_A . We expect trapping to be avoided when the parameter $p_{\text{trap}} \lesssim 1$, with

$$p_{\text{trap}} \cong f_A^{2-m_A} N^2. \quad (10)$$

This result holds independent of subunit concentration and capsid geometry, provided the capsid terminates at a finite size and the subunit-subunit affinity is strong enough to stabilize the capsid. It is important to note that assembly would be sensitive to formation of defective capsids in this limit, which we neglect in this scaling analysis. However, as shown below we do observe this limit in Brownian dynamics simulations where defective assembly is allowed.

IV. NUMERICAL RESULTS

A. Effect of Allostery on Assembly Robustness

In this section, we test the scaling predictions against results from the Brownian dynamics (BD) simulations and master equation model. For the BD simulations, we focus on quantities involving variation of the subunit binding energy parameter ϵ_b , as varying this parameter is more computationally tractable than varying subunit concentrations over the range needed to test scaling. Specifically, we test how the assembly

yield and median assembly time depend on ϵ_b and the activation energy g_A , and we test the prediction that a sufficiently large value of g_A enables assembly even in the limit of high subunit-subunit binding affinity. The latter test is particularly important since the BD simulations allow for the formation of defective capsids. We then vary concentration using the master equation model.

We begin by examining the time dependence of the fraction of subunits in complete capsids (P_N) as the binding energy is varied. A capsid is defined as a structure containing 20 subunits, each with strong interactions with three neighbors. This definition counts only configurations which correspond to small fluctuations around the icosahedral ground state. We note that experimental measurements of assembly kinetics commonly employ SEC, which monitors P_N , or light scattering, which under certain conditions monitors the mass-averaged molecular weight of assemblies [11]. Ref. [66] showed that P_N and assembly molecular weight closely track each other below the crossover concentration c_c (Eq. (4)), and quantities such as the lag phase and nucleation time follow the same scaling laws when calculated from either observable. Therefore, here we present only P_N .

Fig. 3A shows P_N as a function of time for several values of ϵ_b with the ‘No Allostery’ model. We see that initially as ϵ_b increases the lag phase shortens, P_N rises in time more rapidly, and asymptotes at a higher value. However, at the highest ϵ_b shown, $\epsilon_b = 30$, the fraction capsid quickly saturates at a low value, indicative of kinetic trapping. To illustrate this point, Fig. 3B shows $P_N(t)$ for two values of ϵ_b , overlaid with a plot of the fraction of subunits in intermediates, $P_{\text{int}}(t) = \sum_{n=2}^{N-1} n c_n(t) / c_0$. We see that for $\epsilon_b = 14$ the intermediates peak near the end of the lag phase, and then rapidly fall as capsids are produced. In contrast, for $\epsilon_b = 30$, nearly all subunits are trapped in intermediates. Thus capsids are only slowly produced when larger intermediates scavenge subunits from smaller ones. Since we find that it correlates well with other measures of kinetic trapping, we use P_{int} to characterize trapping as a function of parameter values.

For comparison, Figs. 3C,D show the time-dependence of P_N and P_{int} calculated from the master equation with varying subunit concentration [82]. We see that the kinetics and the relationship between P_{int} and the onset of kinetic trapping are consistent with the BD results.

Next, we consider how assembly robustness changes when allostery is introduced. Fig. 4A shows P_N measured at long but finite times as a function of ϵ_b for the three conformation cases with $g_A = 4$. We see that in both cases, allostery shifts the onset of assembly to dramatically higher binding affinity. More significantly, assembly stays remarkably productive up to extremely high values of ϵ_b for both allostery mechanisms. This observation corresponds to the high affinity limit discussed above (Eq. (10)). Analysis of simulation trajectories shows that the eventual decline in P_N at high ϵ_b arises due to the formation of defective capsids.

To show how assembly robustness depends on the strength of allostery, we compare P_N calculated for the ‘No Allostery’ with ‘Induced Fit’ at two values of g_A in Fig. 4B. To aid in comparing these cases, we have ‘undone’ the shift in the bind-

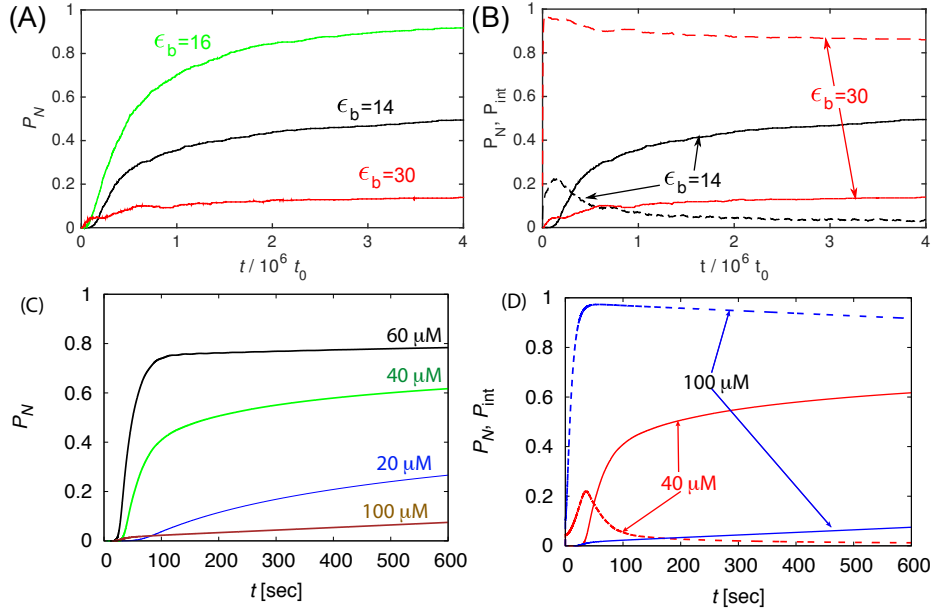


FIG. 3. The dependence of assembly kinetics on parameter values for ‘No Allostery’. (A) The fraction of subunits in complete capsids P_N observed in Brownian dynamic simulations (BD) is shown as a function of time for indicated values of the binding energy parameter ϵ_b . (B) The fraction capsid (solid lines) is compared to the fraction of subunits in intermediates P_{int} (dashed lines) for small and large values of ϵ_b . For (A) and (B), each simulation is run until $1.5 \times 10^7 t_0$, and each data point corresponds to an average over 20 independent simulations. (C) The fraction capsid as a function of time measured for the master equation is shown at indicated total subunit concentrations c_0 , with binding affinity $g_b = -7$ (all energies are in units of $k_B T$). (D) The fraction capsid is compared to the intermediate fraction for the master equation, for concentrations below and above the trapping point $c_{\text{kt}} = 60 \mu\text{M}$ (Eq. 5).

ing energies suggested by Eq. (7) according to

$$\epsilon_b^{\text{eff}} = \epsilon_b + \frac{\hat{n} + 1}{\hat{n} - 1} \log f_A, \quad (11)$$

so that the onset of productive assembly occurs roughly at the same value of ϵ_b^{eff} for each case. We note that the curves do not line up perfectly because this mapping is only approximate, since the free energy includes binding entropy factors whose magnitude depend logarithmically on ϵ_b (see Eq. (12) in the appendix) and the degeneracy of available binding sites on a given structure [71, 83, 84]. We see that the smaller activation energy $g_A = 2$ leads to assembly over approximately the same range of ϵ_b^{eff} as for the ‘No Allostery’ case, supporting our earlier conclusion that, below the high affinity limit, allostery does not significantly increase assembly robustness.

To further understand the onset of the high affinity limit and robust assembly, we examine the dependence of kinetic trapping on parameter values. First, to support the earlier statement that P_{int} is a good metric for the extent of kinetic trapping, Fig. 5 compares P_N and the intermediate fraction P_{int} measured at long times for ‘No Allostery’; we see that a rise in P_{int} correlates with the decline in P_N .

Next we focus on high binding affinity, $\epsilon_b = 30$, and show P_{int} in Fig. 6A for different values of g_A , plotted against the trapping parameter p_{trap} (Eq. (10)). We see that for ‘No Induced Fit’ the onset of trapping occurs when p_{trap} is of order one as expected, but trapping for ‘Induced Fit’ is shifted to lower values of p_{trap} . Analysis of simulation trajectories sug-

gests that this difference arises due to the formation of malformed capsids, which are more prevalent in the ‘Induced Fit’ trajectories. We can understand this observation by noting that the probability of a defective capsid increases with the rate of subunit addition, which is larger for ‘Induced Fit’ (since inactive subunits can bind to growing capsids, and because a given value of p_{trap} corresponds to higher f_A for ‘Induced Fit’ in comparison to ‘No Induced Fit’). As further evidence for the importance of defective capsids, Fig. 6B shows the equivalent plot from the rate equation model, which does not allow for malformed capsids. In that case, we see that results from both allostery mechanisms fall approximately on the same curve. We have also tested the predicted scaling against capsid size N in that figure by including results from rate equation calculations with N varying from 20 to 240.

Finally, we consider the effect of allostery on assembly timescales. Fig. 7A shows the median assembly time $\tau_{1/2}$ (defined by $P_N(\tau_{1/2}) = 1/2$) calculated from the master equation as a function of initial subunit concentration for each conformation case, with modest interaction parameter values ($g_b = -7$, $g_A = 2$). This plot allows testing of several scaling predictions. Firstly, we see that both in the presence and absence of allostery, the $\tau_{1/2}$ at low concentrations scales as $c_0^{n_{\text{nuc}}-1}$ as predicted by Eq. (3). As expected, median times are equal for ‘No Induced Fit’ and ‘Induced Fit’ at low concentrations where the reaction is nucleation-limited, since the autostery does not affect nucleation in our model. Secondly, the dashed vertical lines show the point of kinetic trapping c_{kt}

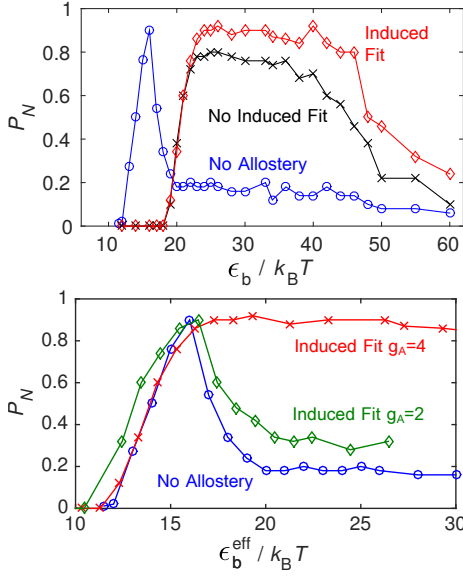


FIG. 4. (top) Capsid fraction P_N as a function of the binding energy parameter ϵ_b in Brownian dynamics (BD) simulations for the three conformation cases, with $g_A = 4k_B T$. (bottom) The capsid fraction is shown as a function of the shifted binding energy ϵ_b^{eff} (Eq. 11) for ‘No Allostery’ and ‘Induced Fit’ with two activation energy values. For both panels, other parameters are as in Fig. 3.

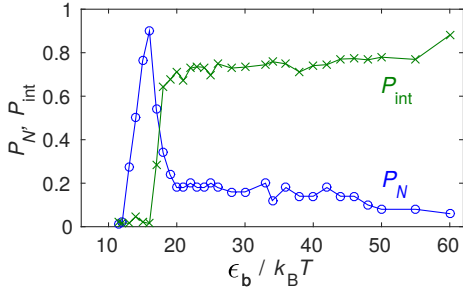


FIG. 5. Capsid fraction P_N and intermediate fraction P_{int} as a function of the binding energy ϵ_b in Brownian dynamics (BD) simulations for ‘No Allostery’.

predicted by Eq. (5) for each case. We see that the median assembly time takes off above these threshold concentrations, indicating the presence of trapping.

Notice that even a modest conformation energy $g_A = 2$ shifts the region of productive assembly to unrealistically high subunit concentrations, due to the scaling of nucleation times with f_A . However, this effect can be counteracted by increasing the subunit binding affinity. To illustrate this point, and to test the predicted effect of conformation specificity on assembly robustness (Eq. (8)), we also calculated median assembly times with the binding affinity for the allostery cases increased according to Eq. (7) (Fig. 7B). As anticipated, the nucleation times become equal at low concentrations, and the ‘Induced Fit’ and ‘No Allostery’ cases enjoy the same range of productive assembly, while the ‘No Induced Fit’ case becomes kinetically trapped at a lower concentration due to its increased elongation time relative to nucleation. To further test the ex-

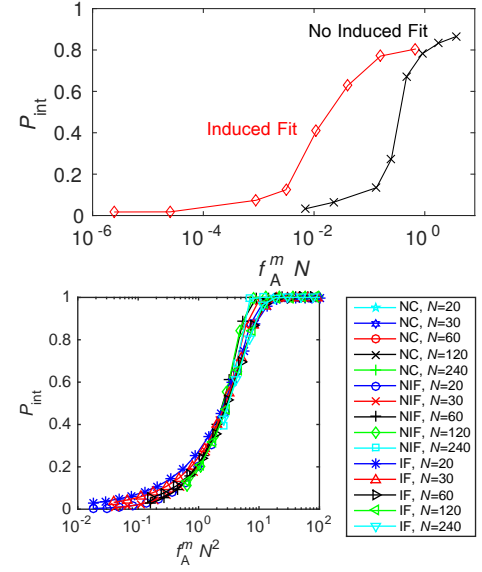


FIG. 6. (top) Kinetic trapping in BD simulations. The magnitude of trapping P_{int} measured in BD simulations is shown in the limit of very high subunit binding strength ($\epsilon_b = 40$). Results are shown as a function of the high affinity trapping parameter $p_{\text{trap}} = f_A^m N^2$, with $m = 1$ or 2 respectively for the ‘No Induced Fit’ and ‘Induced Fit’ cases (see Eq. (10)) and $N = 20$. The parameter f_A was tuned by varying $g_A/k_B T$ from 1 to 6. Other parameters are as in Fig. 3. (bottom) Relationship between capsid size and sensitivity to trapping in the high-affinity limit from the master equation. The magnitude of trapping P_{int} calculated for $g_b = -25$ is shown for the three conformation cases and capsid sizes $N \in [20, 240]$. The parameter f_A was tuned by varying $g_A/k_B T$ from 0 to 10.

pected scaling with activation energy, we show the trapping measure P_{int} as a function of subunit concentration normalized by c_{kt} for all three cases over a range of g_A in Fig. 7C. We see that in all cases trapping takes off as the threshold concentration is crossed.

Fig. 8 shows the variation of $\tau_{1/2}$ with the shifted binding energy for the three allostery cases calculated from BD simulations. We see that the nucleation-dominated regime, marked by an exponential decrease in $\tau_{1/2}$ with increasing ϵ_b^{eff} , roughly overlaps for the three cases, suggesting that the mapping is approximately correct. Notice that with allostery, $\tau_{1/2}$ becomes constant at large ϵ_b^{eff} ; this corresponds to the high affinity limit.

B. Implications of allostery for parameter estimations from experimental data

In this section we compare assembly kinetics with and without allostery, and consider ramifications for parameter estimations made using models that do not account for allostery. Fig. 9 compares the time-dependence of P_N at several concentrations for the ‘No Allostery’ and ‘Induced Fit’ cases calculated from the master equation, with the subunit binding affinity in the ‘Induced Fit’ case adjusted according to Eq. (7). In particular, with $g_b = -7$ for ‘No Allostery’, we obtain $g_b = -10.54$ for ‘Induced Fit’ with $g_A = 2$. We see

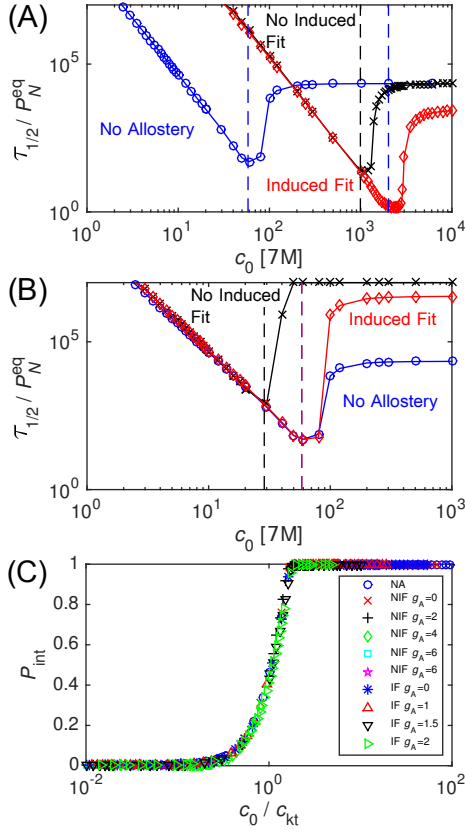


FIG. 7. The effect of allostery on assembly timescales calculated from the master equation. **(A)** The median assembly time $\tau_{1/2}$ is shown as a function of initial subunit concentration for the three allostery cases, with the subunit binding affinity $g_b = -7$ and the activation energy $g_A = -2$. For each case, the vertical dashed line indicates the point of kinetic trapping c_{kt} calculated by Eq. (5). **(B)** The median assembly time is shown as a function of concentration for the three cases, but with the subunit binding affinity shifted according to Eq. (7), $g_b = -7$ for NC and $g_b = -10.54$ for NA and A, with $g_A = 2$. The vertical lines indicate the predicted location of c_{kt} for these modified subunit affinities. **(C)** The maximum fraction of subunits found in intermediates, P_{int} , is shown as a function of initial subunit concentration normalized by the trapping threshold c_{kt} (Eq. 5), for indicated allostery cases and values of the activation energy g_A .

that this relationship, derived to relate assembly kinetics in the nucleation-dominated regime, leads to assembly kinetics which also closely match throughout early times.

The most significant difference between the kinetics with and without allostery appears at a very early times. As the kinetics transition from the lag phase to the rapid production of capsids, the takeoff is more rapid for the ‘Induced Fit’ case. A similar observation was made by Chen et al. [14] and shown to be consistent with light scattering from assembling BMV viruses. This observation can be understood by noting that, due to the stronger binding affinity in the ‘Induced Fit’ case, kinetics in the elongation phase are closer to irreversible and thus the distribution of lag times is more sharply peaked.

Despite its modest impact on the form of kinetics, allostery

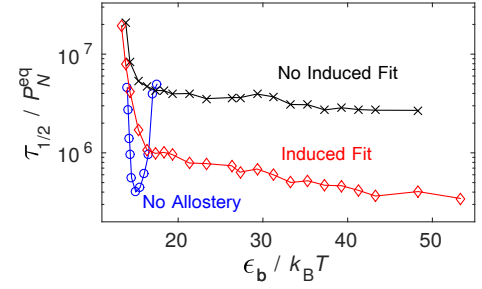


FIG. 8. The median assembly time $\tau_{1/2}$ measured in BD simulations is shown for the three conformation cases as a function of ϵ_b^{eff} , with $g_A = 4$. Other parameters are as in Fig. 3 except simulations were run until $2.6 \times 10^7 t_0$.

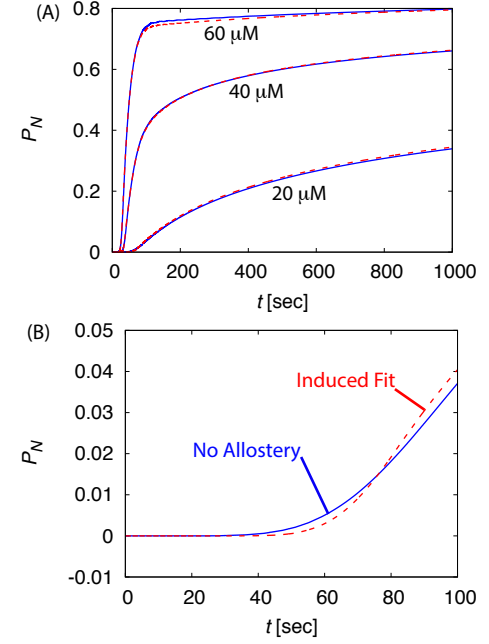


FIG. 9. The time-dependence of assembly is compared between the ‘No Allostery’ (solid lines) and ‘Induced Fit’ (dashed lines) cases for $c_0 = 20 \mu\text{M}$. **(A)** The fraction capsid is shown for indicated initial subunit concentrations. **(B)** The early time course of assembly is shown for both cases at $c_0 = 20 \mu\text{M}$, showing that ‘Induced Fit’ exhibits a sharper take off at the end of the lag phase in comparison to ‘No Allostery’.

can dramatically skew quantitative parameter values estimated from experimental data. Fig. 9 demonstrates that a fit against assembly kinetics not accounting for allostery would underestimate g_b , by an amount proportional to g_A (see Eq. (7)). We emphasize that the shift factor in g_b^{shift} is independent of subunit concentration. Thus, testing data fits against multiple subunit concentrations does not necessarily identify the presence of allostery. A binding affinity underestimated at one subunit concentration (due to not accounting for allostery) would be consistent with data at other subunit concentrations.

A second approach to estimate subunit binding affinities is to measure the ratio of capsid to free subunits at very long

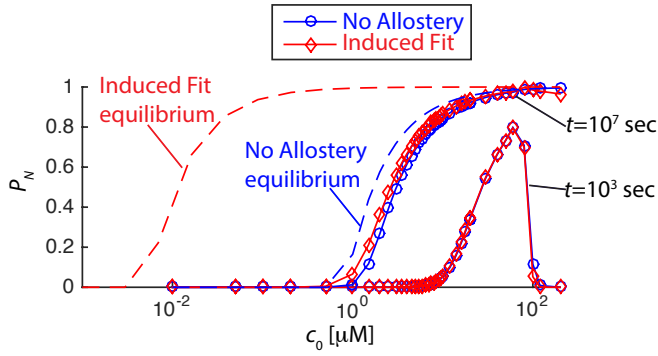


FIG. 10. The difficulties of detecting allostery from finite-time experiments. The fraction of subunits in capsids calculated from the master equation as a function of total subunit concentration is shown for 10^3 and 10^7 seconds, for ‘No Allostery’ (o symbols) and ‘Induced Fit’ (◊ symbols). The equilibrium values of P_N are shown for each case as dashed lines. Parameters are $g_b = -7$ for ‘No Allostery’ and $g_b = -10.54$ and $g_A = 2$ for ‘No Allostery’.

times, and fit the results to the equilibrium law of mass action [28]. As has been pointed out previously [2, 45, 66], P_N only asymptotically approaches its equilibrium value, and thus fits assuming equilibrium will tend to underestimate g_b . We find that unlooked for allostery can dramatically enhance underestimation, since the lack of distinguishability seen in Fig. 9 persists for very long times. In Fig. 10 we show the fraction capsid P_N as a function of subunit concentration predicted from the master equation with and without allostery, for the same interaction parameters as in Fig. 9. We see that the curves overlap perfectly at 1000 seconds, and still nearly overlap at 10^7 seconds (about 4 months). By this time, the ‘No Allostery’ system has nearly reached its equilibrium, and thus the equilibrium subunit binding affinity estimated from these results would be reasonably accurate. However, for ‘Induced Fit’ the binding affinity is underestimated by $3.5k_B T$. Larger values of g_A lead to more severe underestimation with similar lack of distinguishability between the two cases.

V. DISCUSSION AND OUTLOOK

It has been well established that the assembly of rigid subunits into ordered structures requires weak, reversible interactions, as interactions which are strong in comparison to the thermal energy lead to kinetic traps [1–3, 71, 74]. Here, we have used computational and theoretical models to investigate how this constraint changes when the subunits have internal degrees of freedom, such as conformational states, allowing them to change their capacity for interaction during the assembly process.

We find that a sufficient bias in the free subunit population toward the inactive conformation allows productive assembly at very high subunit concentrations or binding affinities. In particular, allostery differentially regulates the rates of nucleation and elongation, thus suppressing the kinetic trap that

arises when free subunits are depleted before capsid elongation is completed. To our knowledge, such robust assembly has not been observed in vitro, but this effect could be important for reactions such as the assembly of the mature HIV capsid which occurs at high concentration within the budded viral particle [85]. However, this mechanism does not provide complete protection against kinetic traps that arise due to defective capsid assembly, as we observed in the Brownian dynamics simulations at very high ϵ_b . The prevalence of such defects, and hence the actual range available for productive assembly, will depend on the orientational specificity of the subunit-subunit interactions [71, 84], which has not yet been evaluated for specific capsid proteins.

Given that increasing g_A can qualitatively change assembly robustness, it would be of interest to measure the subunit conformational equilibrium as a function of protein sequence and solution conditions. We have recently used all-atom simulations to estimate a free energy difference of about $3k_B T$ between two quasi-equivalent conformations (meaning different conformations found at positions with different local symmetry in the capsid) of the MS2 coat protein. However, we note that the accuracy of such calculations is necessarily limited by force field accuracy and the quality of sampling. The simulations also found that RNA binding could significantly shift the populations. While similar calculations are possible for the active/inactive transition, they will be complicated by the fact that the inactive ‘conformation’ can be an ensemble with significant structural diversity [9].

Our findings have several implications for interpreting mechanisms and estimating interaction parameters from experimental data. Firstly, we find that the kinetics of an assembly reaction with allostery are quite difficult to distinguish from those of a reaction with no conformation dependence at moderate parameters (Fig. 9). One commonly used test for the quality of a model is to fit interaction parameters to kinetics at one subunit concentration, and then test their predictions against kinetics measured at other concentrations. However, the binding affinity adjustment Eq. (7) used to match parameters between the ‘No Allostery’ and ‘Induced Fit’ cases in Fig. 9 is independent of concentration, and we find that the kinetics match quite closely even into the regime where kinetic trapping starts to set in. Thus, a binding affinity underestimated at one subunit concentration (due to not accounting for allostery) would also appear consistent with data at other subunit concentrations. While it might be argued that such a close match arises because of the simplicity of the rate equation model, parameter estimation from experimental data generally relies on models with similar levels of approximations to enable computational tractability. Moreover, we observe a similar matching of kinetics in the BD simulations. Thus, our results suggest that strong emphasis should be placed on matching the very early phases of assembly kinetics during parameter estimation. The importance of fitting the lag phase was also suggested by Chen et al. [14], who furthermore demonstrated this phase can be monitored by light scattering with millisecond resolution. Techniques sensitive to individual capsids (e.g. [26]) will allow further investigation of early-time kinetics.

As we note above, by shifting productive assembly to higher binding affinities, allostery increases the maximum capsid thermostability that is kinetically accessible. Thus, allostery, in the form of conformational transitions during assembly, may offer an alternative strategy to post-assembly conformational changes or covalent modifications used by some bacteriophages to stabilize their capsids (*e.g.* [86, 87]).

Finally, there are several effects we did not investigate here which could lead to additional control over assembly. We have focused on the limit in which subunit conformational transitions are fast in comparison to assembly timescales. Our approaches are easily generalized to other conformational timescales. A preliminary investigation showed that decreasing the conformational transition rate allows higher assembly yields in the high affinity limit for the ‘Induced Fit’ case (similar to the case of bacterial flagella assembly [51]). While we focus here on allostery at the level of protein-protein interactions, there are many examples in the literature suggesting that interaction with non-protein components, such as RNA, lipid membranes, or small molecule assembly effectors can exert additional allosteric control on protein conformations (see [6]). Understanding how these multiple regulatory mechanisms cooperate to control the time, place, and rate of assembly will lead to a more complete understanding of viral life cycles, and also may identify new strategies for designing human-made assembly systems.

VI. APPENDIX

A. Comparison between Computational and Master Equation Models

In the main text our analysis of the master equation focuses on capsid sizes and parameters consistent with experiments on HBV [11]. Here we consider parameters roughly corresponding to those of the BD simulations to further evaluate the degree of similarity between master equation and BD results. To minimize reliance on data fitting, we calculated the subunit-subunit binding free energy g_b as a function of the simulation well-depth ϵ_b according to $g_b(n) = -\Delta n_{c,n} \epsilon_b - T(s_b + s_c(n))$ where $\Delta n_{c,n}$ is the number of subunit-subunit contacts added to form a cluster of size n (assuming the lowest energy configuration at each size n), $s_b(n)$ denotes the translational and rotational binding entropy penalties and s_c denotes the change in ‘configurational’ entropy associated with degeneracy of the lowest energy configuration for each cluster size n . The binding entropy can be estimated from a saddle point approximation of the partition function for a subunit dimer as [71, 83]

$$s_b/k_B \approx -\frac{3}{2} \log \left. \frac{\beta \partial^2 U_{\text{att}}(r)}{\partial r^2} \right|_{r=2^{1/6}\sigma} - \frac{1}{2} \log \frac{(\beta \epsilon_b)^3 \pi^7}{\theta_c^4 \phi_c^2}. \quad (12)$$

We note that this approximation provides an accurate description of the dimerization free energy but neglects additional entropy penalties incurred by subunits forming more than one contact. The number of contacts $\{\Delta c(n)\}$ and configurational entropy $\{s_c(n)\}$ values for building an icosahedron are

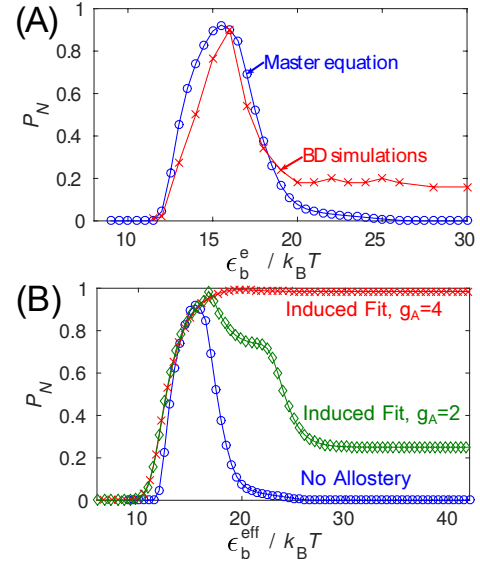


FIG. 11. (A) The fraction capsid P_N predicted by BD simulations and the master equation as a function of the binding energy parameter ϵ_b for ‘No Allostery’, for $t = 10^7 t_0$. Parameters for the BD simulations are as in Fig. 3, and the master equation parameters are set correspondingly (see the text in Appendix VI A). (B) The master equation results for fraction capsid are shown as a function of the shifted binding energy ϵ_b^{eff} (Eq. 7) for ‘No Allostery’ and ‘Induced Fit’ with two activation energy values (compare to the bottom panel of Fig. 4). The only adjustable parameter for the master equation in these results is the association rate constant k , which was set (by eye) to $k = 0.003\sigma^3/t_0$.

given in Table S2 of Roldao et al. [88]. The shifted binding energy (Eq. (11)) is modified to $\epsilon_b^{\text{eff}} = \epsilon_b + \frac{\bar{n}+1}{\bar{n}-1} \log f_A + 3 \log (\epsilon_b^{\text{eff}}/\epsilon_b)$.

The subunit association rate k is the only adjustable parameter we used in this comparison. While k could be estimated directly from simulations, its value changes for each cluster size due to different occluded volumes neighboring subunits. For simplicity, we use one average value $k = 0.003\sigma^3/t_0$, which we estimated (by eye) by comparing master equation and simulation results for $P_N(t)$ for several values of ϵ_b .

Although the master equation kinetics do not perfectly match BD results, their agreement is reasonable considering the approximations in our estimate of g_b and the simplifications inherent in the master equation. For example, in Fig. 11 we show the master equation results for P_N calculated at long time $10^7 t_0$ as a function of the binding energy parameter and activation energy. The most significant difference is that the BD simulations result in a small number of completed capsids at large ϵ_b even for ‘No Allostery’, whereas the master equation does not. We believe this difference arises due to binding of oligomers in the BD simulations [71] which is not accounted for in the master equation. We note however that the effect of oligomer binding would diminish at smaller subunit concentrations.

B. Rate equation model with conformation changes

Extension of Eqs. (1) to include interconversion of free subunits between inactive and active conformations results in

$$\frac{dc_1}{dt} = -k_1 c_1^2 + 2\bar{k}_2 c_2 + \sum_{n=2}^N -k_n c_n c_1 + \bar{k}_n c_n \quad (13)$$

$$+ k_A c_1 - \bar{k}_A c_1$$

$$\frac{dc_n}{dt} = (k_{n-1} c_1 + k_{n-1}^* c_1^*) c_{n-1} - (k_n c_1 + k_n^* c_1^*) c_n$$

$$- (\bar{k}_n + \bar{k}_n^*) c_n + (\bar{k}_{n+1} + \bar{k}_{n+1}^*) c_{n+1}$$

for $n = 2 \dots N$ (14)

$$\frac{dc_1^*}{dt} = \sum_{n=2}^N -k_n^* c_n c_1^* + \bar{k}_n^* c_n - k_A c_1^* + \bar{k}_A c_1 \quad (15)$$

where c_1^* and c_1 are respectively the concentrations of inactive and active free subunits, k_A and \bar{k}_A are the active/inactive interconversion rate constants related by $k_A/\bar{k}_A = \exp(-g_A/k_B T)$, and k_n^* and \bar{k}_n^* are the rate constants for association and dissociation of inactive subunits to nucleated partial capsids. For simplicity, we take the nucleus size for inactive subunits binding equal to the energetic critical nucleus size, and the association rate constant for inactive and active subunits to be equal above this size. Specifically, for ‘Induced Fit’, $k_n^* = \Theta(n - n_{\text{nuc}}) f_n$ with $\Theta(n)$ the Heaviside function and f_n the association rate constant for active subunits, while for ‘No Induced Fit’, $k_n^* = 0 \forall n$. Alternate choices for these quantities do not qualitatively affect the results. Finally, dissociation rate constants are given by detailed balance, $\bar{k}_n^* = k_n^* \exp[(\Delta g_n - g_A)/k_B T]/v_0$, where the free energy change upon association of an inactive subunit includes the activation energy g_A . The initial condition is $c^*(0) = (1 - f_A) c_0$, $c_1(0) = f_A c_0$, $c_n(0) = 0$ for $n > 1$.

C. Brownian dynamics model details

We use the patchy-sphere model presented in Ref. [68]; our description here closely follows that reference. The minimum energy structure is a complete capsid of 20 subunits, which have a spherical excluded volume with three attractive patches, or bond vectors, that are separated by 108° and rotate rigidly with the subunit. The attractive interaction between two complementary bond vectors on respective subunits i and j is maximized when (1) the distance between the attractors r_{ij}^b is minimized, (2) the angle θ_{ij}^b between bond vectors is minimized, and (3) the dihedral angle ϕ_{ij}^b calculated from two secondary bond vectors, which are not involved in the primary interaction, is minimized. A schematic of the subunit interactions is shown in Fig. 2. Minimizing ϕ_{ij}^b creates an interaction that resists torsion and enforces angular specificity commensurate with a complete capsid. The potentials are given by Eqs (16) [68, 89]

$$U = U_{\text{rep}}(R_{ij}) + \sum_b U_{\text{att}}(r_{ij}^b) S(\theta_{ij}^b, \phi_{ij}^b)$$

$$U_{\text{rep}}(R_{ij}) = \mathcal{L}_{12}(R_{ij}, 2^{\frac{1}{6}} \sigma, \sigma)$$

$$U_{\text{att}}(r_{ij}) = \epsilon_b \mathcal{L}_{12}((r_{ij} + 2^{\frac{1}{6}} \sigma), 2^{\frac{1}{6}} \sigma, r_c)$$

$$S(\theta, \phi) = \frac{1}{4} \Theta(\theta - \theta_c) \Theta(\phi - \phi_c)$$

$$(\cos(\pi\theta/\theta_c) + 1) (\cos(\pi\phi/\phi_c) + 1) \quad (16)$$

with \mathcal{L}_p a generalized truncated and shifted Lennard-Jones function:

$$\mathcal{L}_p(x, x_c, \sigma) \equiv 4 \left(\left(\frac{x}{\sigma} \right)^{-p} - \left(\frac{x}{\sigma} \right)^{-p/2} - \left(\frac{x_c}{\sigma} \right)^{-p} + \left(\frac{x_c}{\sigma} \right)^{-p/2} \right) \Theta(x - x_c) \quad (17)$$

In Eq. (16) the index b sums over pairs of complementary bond vectors, $\Theta(x)$ is the Heaviside step function and R_{ij} is the subunit center-to-center distance.

Conformational dynamics. In our BD simulations with conformational dynamics, free subunits stochastically switch between inactive and active conformations. The only difference between the two conformations is their interaction partners. For ‘No Induced Fit’, inactive subunits do not experience attractive interactions with any subunit (but they still experience the repulsive excluded volume with all types of subunits), while pairs of active subunits experience the attractive interactions described above. ‘Induced Fit’ has a similar matrix of interactions, except that in active subunits experience attractive interactions with any subunit in a partial capsid which has at least one completed polygon (meaning a closed cycle of interactions among subunits). In particular, inactive subunits experience no attractions with other inactive subunits, free active subunits, or subunits in partial capsids with no completed polygon, but do experience attractions to partial capsids with at least one completed polygon. Pairs of active subunits experience attractions as described above. Since we focus on the limit of fast conformational dynamics (relative to assembly timescales), subunits underwent conformational sampling with frequency $2.5t_0^{-1}$. At this frequency, the conformation of each free subunit was stochastically set to inactive or active with respective probabilities $1 - f_A$ and f_A .

Acknowledgements

This work was supported by the NIH, Award Number R01GM108021 from the National Institute Of General Medical Sciences and the Brandeis Center for Bioinspired Soft Materials, an NSF MRSEC, DMR-1420382. Computational resources were provided by the NSF through XSEDE computing resources and the Brandeis HPCC which is partially supported by the Brandeis MRSEC.

- [1] S. Whitelam and R. L. Jack, *Ann Rev Phys Chem* **66**, 143 (2015).
- [2] M. F. Hagan, *Adv. Chem. Phys.* **155**, 1 (2014).
- [3] A. Zlotnick, *Virology* **315**, 269 (2003).
- [4] A. Rao, *Annu. Rev. Phytopathol.* **44**, 61 (2006).
- [5] R. F. Garmann, A. Gopal, S. S. Athavale, C. M. Knobler, W. M. Gelbart, and S. C. Harvey, *RNA* **21**, 877 (2015).
- [6] J. D. Perlmutter and M. F. Hagan, *Annu. Rev. Phys. Chem.* **66**, 217 (2015), <http://dx.doi.org/10.1146/annurev-physchem-040214-121637>.
- [7] P. G. Stockley, N. A. Ranson, and R. Twarock, *Future Virology* **8**, 531 (2013).
- [8] D. L. D. Caspar, T. Rosenstiel, and B. Medical, *Biophys. J.* **32**, 103 (1980).
- [9] C. Packianathan, S. P. Katen, C. E. Dann, and A. Zlotnick, *J. Virol.* **84**, 1607 (2010).
- [10] P. E. Prevelige, D. Thomas, and J. King, *Biophys. J.* **64**, 824 (1993).
- [11] A. Zlotnick, J. M. Johnson, P. W. Wingfield, S. J. Stahl, and D. Endres, *Biochemistry* **38**, 14644 (1999).
- [12] A. Zlotnick, R. Aldrich, J. M. Johnson, P. Ceres, and M. J. Young, *Virology* **277**, 450 (2000).
- [13] G. L. Casini, D. Graham, D. Heine, R. L. Garcea, and D. T. Wu, *Virology* **325**, 320 (2004).
- [14] C. Chen, C. C. Kao, and B. Dragnea, *J. Phys. Chem. A* **112**, 9405 (2008).
- [15] C. Berthet-Colominas, M. Cuillel, M. H. J. Koch, P. Vachette, and B. Jacrot, *Eur. Biophys. J. with Biophys. Lett.* **15**, 159 (1987).
- [16] S. Kler, R. Asor, C. Li, A. Ginsburg, D. Harries, A. Oppenheim, A. Zlotnick, and U. Raviv, *J. Am. Chem. Soc.* **134**, 8823 (2012).
- [17] S. Kler, J. C.-Y. Wang, M. Dhason, A. Oppenheim, and A. Zlotnick, *ACS Chem. Biol.* **8**, 2753 (2013).
- [18] P. G. Stockley, O. Rolfsson, G. S. Thompson, G. Basnak, S. Francese, N. J. Stonehouse, S. W. Homans, and A. E. Ashcroft, *J. Mol. Biol.* **369**, 541 (2007).
- [19] G. Basnak, V. L. Morton, O. Rolfsson, N. J. Stonehouse, A. E. Ashcroft, and P. G. Stockley, *J. Mol. Biol.* **395**, 924 (2010).
- [20] C. Uetrecht, I. M. Barbu, G. K. Shoemaker, E. van Duijn, and J. R. Heck Albert, *Nat Chem* **3**, 126 (2011).
- [21] V. Baumgaertel, B. Mueller, and D. C. Lamb, *Viruses* **4**, 777 (2012).
- [22] N. Jouvenet, S. M. Simon, and P. D. Bieniasz, *J. Mol. Biol.* **410**, 501 (2011).
- [23] A. Borodavka, R. Tuma, and P. G. Stockley, *Proc. Natl. Acad. Sci. U. S. A.* **109**, 15769 (2012).
- [24] G. Tresset, C. Le Coeur, J.-F. Bryche, M. Tatou, M. Zeghal, A. Charpilienne, D. Poncet, D. Constantin, and S. Bressanelli, *J. Am. Chem. Soc.* **135**, 15373 (2013).
- [25] E. E. Pierson, D. Z. Keifer, L. Selzer, L. S. Lee, N. C. Contino, J. C. Y. Wang, A. Zlotnick, and M. F. Jarrold, *J. Am. Chem. Soc.* **136**, 3536 (2014).
- [26] K. Zhou, L. Li, Z. Tan, A. Zlotnick, and S. C. Jacobson, *J. Am. Chem. Soc.* **133**, 1618 (2011).
- [27] Law-Hine, D. and Sahoo, A. K. and Bailleux, V. and Zeghal, M. and Prevost, S. and Maiti, P. K. and Bressanelli, S. and Constantin, D. and Tresset, G., *J. Phys. Chem. Lett.* **6**, 3471 (2015).
- [28] P. Ceres and A. Zlotnick, *Biochemistry* **41**, 11525 (2002).
- [29] A. Zlotnick, J. Z. Porterfield, and J. C.-Y. Wang, *Biophys. J.* **104**, 1595 (2013).
- [30] A. Zlotnick, *J. Mol. Biol.* **241**, 59 (1994).
- [31] D. Endres and A. Zlotnick, *Biophys. J.* **83**, 1217 (2002).
- [32] P. Moisan, H. Neeman, and A. Zlotnick, *Biophys. J.* **99**, 1350 (2010).
- [33] P. van der Schoot and R. Zandi, *Phys. Biol.* **4**, 296 (2007).
- [34] Q. Zhang, S. Gupta, T. Emrick, and T. Russell, *J. Am. Chem. Soc.* **128**, 3898 (2006).
- [35] T. Keef, C. Micheletti, and R. Twarock, *J. Theor. Biol.* **242**, 713 (2006).
- [36] M. Hemberg, S. Yaliraki, and M. Barahona, *Biophys. J.* **90**, 3029 (2006).
- [37] E. C. Dykeman, P. G. Stockley, and R. Twarock, *Phys Rev E* **87**, 022717 (2013).
- [38] G. R. Smith, L. Xie, B. Lee, and R. Schwartz, *Biophys. J.* **106**, 310 (2014).
- [39] F. Jamalyaria, R. Rohlf, and R. Schwartz, *J. Comput. Phys.* **204**, 100 (2005).
- [40] B. Sweeney, T. Zhang, and R. Schwartz, *Biophys. J.* **94**, 772 (2008).
- [41] A. Zlotnick and S. Mukhopadhyay, *Trends Microbiol.* **19**, 14 (2011).
- [42] M. S. Kumar and R. Schwartz, *Phys. Biol.* **7**, 045005 (2010).
- [43] L. Xie, G. Smith, X. Feng, and R. Schwartz, *Biophys. J.* **103**, 1545 (2012).
- [44] S. Singh and A. Zlotnick, *J. Biol. Chem.* **278**, 18249 (2003).
- [45] W. H. Roos, R. Bruinsma, and G. J. L. Wuite, *Nat. Phys.* **6**, 733 (2010).
- [46] C. Uetrecht, N. R. Watts, S. J. Stahl, P. T. Wingfield, A. C. Steven, and A. J. R. Heck, *Phys. Chem. Chem. Phys.* **12**, 13368 (2010).
- [47] L. Deshmukh, C. D. Schwieters, A. Grishaev, R. Ghirlando, J. L. Baber, and G. M. Clore, *J. Am. Chem. Soc.* **135**, 16133 (2013), pMID: 24066695, <http://dx.doi.org/10.1021/ja406246z>.
- [48] F. Birnbaum and M. Nassal, *J. Virol.* **64**, 3319 (1990).
- [49] S. J. Stray, C. R. Bourne, S. Punna, W. G. Lewis, M. G. Finn, and A. Zlotnick, *Proc. Natl. Acad. Sci. U. S. A.* **102**, 8138 (2005).
- [50] P. T. Wingfield, S. J. Stahl, R. W. Williams, and A. C. Steven, *Biochemistry* **34**, 4919 (1995).
- [51] S. Asakura, *J. Mol. Biol.* **35**, 237 (1968).
- [52] Y. Kikuchi and J. King, *J. Mol. Biol.* **99**, 695 (1975).
- [53] P. J. G. Butler, *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences* **354**, 537 (1999).
- [54] A. Klug, *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **354**, 531 (1999).
- [55] D. L. Caspar and K. Namba, *Adv. Biophys.* **26**, 157 (1990).
- [56] D. J. Kraft, W. K. Kegel, and P. van der Schoot, *Biophys. J.* **102**, 2845 (2012).
- [57] D. L. Caspar, *Curr. Biol.* **1**, 30 (1991).
- [58] D. L. D. Caspar and A. Klug, *Cold Spring Harbor Symp. Quant. Biol.* **27**, 1 (1962).
- [59] B. Berger, P. W. Shor, L. Tuckerkellogg, and J. King, *Proc. Natl. Acad. Sci. U. S. A.* **91**, 7732 (1994).
- [60] R. Schwartz, P. W. Shor, P. E. Prevelige, and B. Berger, *Biophys. J.* **75**, 2626 (1998).
- [61] R. Schwartz, R. L. Garcea, and B. Berger, *Virology* **268**, 461 (2000).
- [62] O. M. Elrad and M. F. Hagan, *Nano Lett.* **8**, 3850 (2008).
- [63] H. D. Nguyen, V. S. Reddy, and C. L. Brooks, *Nano Lett.* **7**, 338 (2007).

- [64] H. D. Nguyen, V. S. Reddy, and C. L. Brooks, *J. Am. Chem. Soc.* **131**, 2606 (2009).
- [65] V. L. Morton, E. C. Dykeman, N. J. Stonehouse, A. E. Ashcroft, R. Twarock, and P. G. Stockley, *J. Mol. Biol.* **401**, 298 (2010).
- [66] M. F. Hagan, *Biophys. J.* **98**, 1065 (2010).
- [67] R. Zandi, P. van der Schoot, D. Reguera, W. Kegel, and H. Reiss, *Biophys. J.* **90**, 1939 (2006).
- [68] M. R. Perkett and M. F. Hagan, *J. Chem. Phys.* **140**, 214101 (2014).
- [69] A. Branka and D. Heyes, *Phys. Rev. E* **60**, 2381 (1999).
- [70] D. Heyes and A. Branka, *Mol. Phys.* **98**, 1949 (2000).
- [71] M. F. Hagan and D. Chandler, *Biophys. J.* **91**, 42 (2006).
- [72] A. W. Wilber, J. P. K. Doye, A. A. Louis, E. G. Noya, M. A. Miller, and P. Wong, *J. Chem. Phys.* **127**, 085106 (2007).
- [73] S. D. Hicks and C. L. Henley, *Phys. Rev. E* **74**, 031912 (2006).
- [74] D. Rapaport, *Phys. Rev. Lett.* **101**, 186101 (2008).
- [75] D. Endres, M. Miyahara, P. Moisan, and A. Zlotnick, *Protein Sci.* **14**, 1518 (2005).
- [76] J. M. Johnson, J. H. Tang, Y. Nyame, D. Willits, M. J. Young, and A. Zlotnick, *Nano Lett.* **5**, 765 (2005).
- [77] P. Ceres, S. J. Stray, and A. Zlotnick, *J. Virol.* **78**, 9538 (2004).
- [78] A. Zlotnick, *J. Mol. Biol.* **366**, 14 (2007).
- [79] W. K. Kegel and P. van der Schoot, *Biophys. J.* **86**, 3905 (2004).
- [80] This assumption neglects the fact that rotational entropy penalties are most likely not proportional to the number of contacts (see Ref. [83] for further discussion).
- [81] L. Selzer, S. P. Katen, and A. Zlotnick, *Biochemistry* **53**, 5496 (2014), selzer, Lisa Katen, Sarah P. Zlotnick, Adam.
- [82] Since the master equations can be integrated to arbitrarily long times, we report the maximum value of P_{int} observed at any time for a given parameter set.
- [83] M. F. Hagan, *J. Chem. Phys.* **130**, 114902 (2009).
- [84] M. F. Hagan, O. M. Elrad, and R. L. Jack, *J. Chem. Phys.* **135**, 104115 (2011).
- [85] W. I. Sundquist and H.-G. Krausslich, *Cold Spring Harbor perspectives in medicine* **2**, a006924 (2012).
- [86] E. R. May, J. Feng, and C. L. Brooks, *Biophys. J.* **102**, 606 (2012).
- [87] D. Veesler and J. E. Johnson, *Annual Review of Biophysics* **41**, 473 (2012), <http://dx.doi.org/10.1146/annurev-biophys-042910-155407>.
- [88] A. Roldao, M. C. M. Mellado, J. C. Lima, M. J. T. Carrondo, P. M. Alves, and R. Oliveira, *Plos Comp Biol* **8**, e1002367 (2012).
- [89] M. F. Hagan, *Phys. Rev. E* **77**, 051904 (2008).